JVI Journals.ASM.org

Structure and Function of AvtR, a Novel Transcriptional Regulator from a Hyperthermophilic Archaeal Lipothrixvirus

N. Peixeiro,[a,e]* J. Keller,[b] B. Collinet,[b]* N. Leulliot,[b]* V. Campanacci,[c] D. Cortez,[a] C. Cambillau,[c] K. R. Nitta,[d]* R. Vincentelli,[c] P. Forterre,[a] D. Prangishvili,[a] G. Sezonov,[a,e] H. van Tilbeurgh[b]

Institut Pasteur, Unité Biologie Moléculaire du Gène chez les Extrémophiles, Département de Microbiologie, Paris, France[a]; Institut de Biochimie et de Biophysique Moléculaire et Cellulaire, Université Paris-Sud, UMR8619-CNRS, Orsay, France[b]; Architecture et Fonction des Macromolécules Biologiques, CNRS, and Université d'Aix-Marseille, UMR 7257, Case 932, Marseille, France[c]; Institut du Biologie de Développement de Marseille Luminy, IBDML, UMR 6216, CNRS, Marseille, France[d]; Université Pierre et Marie Curie, UMR 7138 "Systématique, Adaptation, Evolution," Paris, France[e]

The structural and functional analysis of the protein AvtR encoded by *Acidianus* filamentous virus 6 (AFV6), which infects the archaeal genus *Acidianus*, revealed its unusual structure and involvement in transcriptional regulation of several viral genes. The crystal structure of AvtR (100 amino acids) at 2.6-Å resolution shows that it is constituted of a repeated ribbon-helix-helix (RHH) motif, which is found in a large family of bacterial transcriptional regulators. The known RHH proteins form dimers that interact with DNA using their ribbon to create a central β-sheet. The repeated RHH motifs of AvtR superpose well on such dimers, but its central sheet contains an extra strand, suggesting either conformational changes or a different mode of DNA binding. Systematic evolution of ligands by exponential enrichment (SELEX) experiments combined with systematic mutational and computational analysis of the predicted site revealed 8 potential AvtR targets in the AFV6 genome. Two of these targets were studied in detail, and the complex role of AvtR in the transcriptional regulation of viral genes was established. Repressing transcription from its own gene, *gp29*, AvtR can also act as an activator of another gene, *gp30*. Its binding sites are distant from both genes' TATA boxes, and the mechanism of AvtR-dependent regulation appears to include protein oligomerization starting from the protein's initial binding sites. Many RHH transcriptional regulators of archaeal viruses could share this regulatory mechanism.

Viruses infecting *Archaea*, one of the three domains of life, have been studied for more than 30 years. During that time, more than 50 archaeal viruses have been described, 32 of which infect hyperthermophilic archaea of the kingdom *Crenarchaeota*. Specific studies in different laboratories have revealed that these crenarchaeal viruses possess unique morphologies, distinct from those of viruses infecting bacteria and eukaryotes (1). Consistent with their exceptional morphotypes, more than 90% of the viral genes have no homologues in the sequences available in public databases (2).

In silico analysis has revealed the diversity of transcriptional regulators encoded by different archaeal viruses (2). The most prevalent structural motifs found in these proteins are the helix-turn-helix (HTH) and the ribbon-helix-helix (RHH). Proteins with the latter domain are encoded by nearly all crenarchaeal viruses (2) and some euryarchaeal viruses (3). In addition, some archaeal viruses encode proteins with looped-hinged-helix and Zn finger domains (2). The abundance of genes coding for proteins belonging to the RHH family in the genomes of *Crenarchaea* and their viruses could underline the important role of these proteins in both host and viral gene transcription regulation (4). Despite their abundance in archaeal virus genomes and crucial role in the regulation of viral genome expression, only one archaeoviral transcription factor, protein SvtR encoded by the rudivirus SIRV1, has so far been studied in detail (5). The nuclear magnetic resonance (NMR) structure of the protein revealed a typical RHH fold. The protein was found to form a dimer and bind DNA with its β-sheet face. Two regions within the SIRV1 genome were pinpointed as the SvtR binding sites; the protein was found to act as a repressor of its own gene as well as the gene for the viral structural protein gp30 (ORF1070) (5).

The RHH proteins are well-known transcriptional regulators present in the bacterial and archaeal domains and their respective viruses, where this DNA binding motif appears to be a common structural scaffold (6). Remarkably, RHH proteins seem to be absent in eukaryotes (6). Although more than 4,000 RHH proteins have currently been predicted by *in silico* analysis, only a few have been studied experimentally (6). Since they are widely distributed in these forms of life and because of their predictable importance for the regulation of cellular processes, there has been over the last years a dramatic increase in structural studies of proteins bearing the RHH domain (6).

The structures of 21 RHH proteins are presently available (5, 7–26). Among them, the methionine repressor MetJ (24), the regulator of plasmid copy number CopG (11, 27–29), and the bacte-

Received 29 May 2012 **Accepted** 3 October 2012

Published ahead of print 10 October 2012

Address correspondence to G. Sezonov, guennadi.sezonov@pasteur.fr, or H. van Tilbeurgh, herman.van-tilbeurgh@u-psud.fr.

* Present address: N. Peixeiro, Commissariat à la Énergie Atomique et aux Énergies Alternatives, iBiTec-S, Service de Biologie Intégrative et Génétique Moléculaire, Gif-sur-Yvette, France; B. Collinet, UFR 927 des Sciences de la Vie, Université Pierre et Marie Curie, Paris, France; N. Leulliot, Laboratoire de Cristallographie et RMN Biologiques, UMR CNRS 8015, Université Paris Descartes, Sorbonne Paris Cité, Faculté des Sciences Pharmaceutiques et Biologiques, Paris, France; K. R. Nitta, Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden.

N.P. and J.K. contributed equally to this article.

doi:10.1128/JVI.01306-12

riophage P22 Mnt and Arc repressors (21, 30) could be considered the prototypical RHH proteins. The RHH motif consists of an N-terminal β-strand and two α-helices connected by a short turn. RHH proteins form functional intertwined dimers through β-sheet formation of the N-terminal strands. These dimers present their central two-stranded antiparallel β-sheet to the major groove of cognate DNA, and residues of this sheet are involved in specific DNA base interactions. The second helix of the motif is engaged in nonspecific phosphate backbone interactions. RHH proteins contain either a single RHH domain (as in Arc and CopG) or a RHH repeat (as in TraY), or they can be combined with additional domains (as in NikR, PutA, and ParD). All RHH proteins bind DNA as higher oligomers and contact inverted or tandem repeats within operators.

In this study, we were focused on a 11.98-kDa putative DNA binding protein encoded by *Acidianus* filamentous virus 6 (AFV6_gp29), which we term AvtR (*Acidianus* virus transcriptional regulator). This protein is highly conserved in all members of the *Betalipothrixvirus* genus of the *Lipothrixviridae* family (AFV3, AFV6, AFV7, AFV8, AFV9, and SIFV) (31). All these viruses contain double-stranded DNA (dsDNA) genomes of around 40 kb, and *in silico* analysis revealed a high level of conservation of their organization (31). Sequence analysis of AvtR shows clearly that it has an N-terminal CopG-like RHH motif, but no structural predictions could be made for the C-terminal half of the protein. In order to study proteins involved in crenarchaeal virus replication, we set out to determine the crystal structure of AvtR and to characterize its targets, functions, and mechanism of transcriptional regulation The detailed characterization of this new archaeal regulator will allow a better understanding of the mechanisms of transcription regulation that participate in the control of the virus infection cycle in *Archaea*.

MATERIALS AND METHODS

Cloning, expression, and purification. The coding sequence of AvtR was amplified by PCR using cDNA as a template. The cDNA was cloned in a derived pET9 plasmid with a His tag at the C terminus. Expression was done at 37°C using the *Escherichia coli* Rosetta(DE3)/pLysS strain grown in 2× YT medium (Bio 101 Inc.). When the cell culture reached an optical density at 600 nm (OD_{600}) of 0.8, induction at 37°C was performed for 3 h with 0.5 mM IPTG (isopropyl-β-D-thiogalactopyranoside) (Sigma). Cells were harvested by centrifugation and suspended in buffer A (20 mM Na citrate [pH 5.6], 200 mM NaCl, 5 mM β-mercaptoethanol). Cell lysis was completed by sonication, and the lysate was heated for 20 min at 50°C before centrifugation at 20,000 × *g* for 20 min. The soluble fraction was loaded on an Ni-nitrilotriacetic acid (Ni-NTA) column (Qiagen Inc.) equilibrated with buffer A. The protein was eluted with imidazole and subsequently loaded on a heparin column (GE Healthcare) equilibrated with buffer A′ (20 mM Na citrate [pH 5.6], 50 mM NaCl, 10 mM β-mercaptoethanol). Elution was performed using a gradient between buffer A′ and buffer B (20 mM Na citrate, 1 M NaCl, 10 mM β-mercaptoethanol). AvtR was eluted at about 0.7 M NaCl. The His tag was not proteolytically removed before the crystallogenesis experiments. Selenomethionine (SeMet)-substituted AvtR was produced in autoinducible medium at 37°C during 24 h. The medium composition (per liter) was as follows: glycerol, 10 g; glucose, 1 g; lactose, 1 g; KH_2PO_4, 13.6 g; $(NH_4)_2SO_4$, 6.6 g; $Na_2HPO_4 \cdot 2H_2O$, 14.2 g; $MgSO_4 \cdot 7H_2O$, 800 mg; $CaCl_2 \cdot 2H_2O$, 100 mg; thiamine, 4 mg, $FeSO_4 \cdot 7H_2O$, 5 mg; $ZnSO_4 \cdot 7H_2O$, 0.8 mg; $MnSO_4$, 0.8 mg; $CuSO_4 \cdot 5H_2O$, 0.08 mg; $NaMoO_4$, 0.4 mg; H_3BO_3, 1 mg; KI, 0.2 mg; amino acids (L-leucine, L-isoleucine, L-valine, L-tryptophan, L-phenylalanine, L-lysine, L-alanine, L-arginine, L-aspartic acid, glycine, L-asparagine, L-proline, L-serine, L-threonine, L-histidine, L-glutamine, and L-glutamic acid), 200 mg; and selenomethionine, 125 mg. The SeMet-substituted protein was purified as the native protein. The protein sample was judged homogeneous as checked by SDS-PAGE, and the purity was estimated on gel to 98%. After purification, 7 mg and 4 mg of pure protein per liter of culture were obtained for the native protein and the SeMet-substituted protein, respectively.

Structure resolution. Crystals of SeMet-substituted AvtR were grown from a 1:1-μl mixture of protein (8 mg/ml) with 22 to 27% polyethylene glycol 4000 (PEG4000)–0.1 M HEPES (pH 7.5)–5 to 10% isopropanol, using the hanging-drop vapor diffusion method at 23°C. Crystals were soaked in a mixture of mother liquor and 30% glycerol before flash freezing at 100 K. X-ray diffraction data were collected from a crystal of the SeMet-substituted AvtR on beamline ID29 (ESRF) at the wavelength of the Se K edge. The crystals belong to the $P2_1$ space group with four copies per asymmetric unit, corresponding to a 49.7% solvent content. Data were collected at a resolution of 2.6 Å and processed with the programs MOSFLM (32) and SCALA (33) for merging and scaling. The structure was solved using the SAD method using diffraction data collected at 2.6-Å resolution from SeMet-substituted crystals. Eight selenium atom sites were found with the program PHENIX.HYSS (34) These sites were used for phasing with the program SHARP (35). After solvent flattening with the program RESOLVE (36), the quality of the electron density map allowed automated construction of 80% of the model. The missing residues were built by hand using the O molecular graphics program (37), and the model was refined with REFMAC (38).

Systematic evolution of ligands by exponential enrichment (SELEX) assay. A 56-bp oligonucleotide in which the 16 central bases are completely degenerate [5′-CTCAGGGTCGACTTCAGCG(N_{16})CAACCAGTCGTCGACCAGC-3′] was used. The double-stranded oligonucleotide DNA pool was amplified by PCR using primers 5′-CTCAGGGTCGACTTCAGCG-3′ and 5′-GCTGGTCGACGACTGGTTG-3′. PCR amplification was performed with GoTaq Flexi DNA polymerase (Promega). The PCR products were mixed with AvtR protein in the binding buffer and were incubated at room temperature for 10 min [10 mM Tris-HCl (pH 7.5), 50 mM NaCl, 1 mM $MgCl_2$, 0.5 mM dithiothreitol (DTT), 0.5 mM EDTA, 4% glycerol, 5 μg/ml poly(dI-dC)]. Ni-Sepharose 6 FF resin (GE Healthcare) equilibrated in binding buffer was then added, and the mixture was incubated for an additional 20 min. After washing of nonspecific binding oligonucleotides with binding buffer without poly(dI-dC), bead-protein-DNA complexes were dissolved in Milli-Q water. Retrieved oligonucleotides were amplified by PCR, and the PCR products were used as a DNA pool for the next round of selection. This process was repeated seven times. Selected DNA pools were sequenced by 454 sequencing, and sequences of the central regions of DNA pools were used for further analysis.

Oligonucleotides. Oligonucleotides were purchased from Proligo (Sigma-Aldrich) or obtained by PCR using Phusion high-fidelity DNA polymerase (Finnzymes) as indicated. Double-stranded DNA was obtained by annealing the corresponding single-strand nucleotides following standard techniques. T4 polynucleotide kinase (New England BioLabs) was used for ^{32}P radiolabeling of oligonucleotides used in gel shift experiments.

Single-nucleotide polymorphisms: PAGE-EMSA experiments. PAGE-electrophoretic mobility shift assay (EMSA) analyses were performed using oligonucleotides with specific point mutations on the ATT GTGGTACCACCTT sequence provided by the SELEX. Sixteen different double-stranded 40-bp oligonucleotides with single-nucleotide polymorphisms were constructed as shown in Fig. 2B. [5′-AAAAAAAAAAAAA-(mutated SELEX)-GGGGGGGGGGGG-3′]. Using the mobility of the wild-type sequence as a control, 1.5 pmol of each mutated fragment was incubated with 2 μl of 50% (vol/vol) glycerin and 0.5 μg of salmon sperm in a total of 20 μl of transcription buffer (TB) (50 mM Tris-HCl [pH 8.0], 75 mM KCl, 25 mM $MgCl_2$, and 1 mM dithiothreitol) for 15 min at 68°C with various amounts of purified AvtR as indicated in the figure legends. Five microliters of 5× RB loading dye (10 mM Tris-HCl [pH 7.5], 1 mM

EDTA, 20% glycerol, 100 μg/ml bovine serum albumin [BSA], 1 mg/ml xylene-cyanol) was added. Twenty microliters of the DNA-protein mixtures was deposited in a nondenaturing 8% 37.5:1 acrylamide-bisacrylamide gel. PAGE was run in TBE buffer (89 mM Tris-borate, 2 mM Na EDTA, pH 8.3). After migration, the gel was vacuum dried, exposed with Amersham Biosciences Hyperfilm-MP, and developed with a Kodak X-OMAT 2000 processor. The procedure of vacuum drying, exposure, and development was repeated for all PAGE experiments described below.

***In silico* search for AvtR DNA binding sites.** A PERL script was written, where a weighted matrix was used to search for similar sequences in a given genome. The matrix was built assigning different weights to different nucleotides. The script analyzed both strands of the AFV6 genome using an 18-nucleotide (nt) sliding window. The sequences located within the hypothetical promoter regions, ranging from position −150 to +50 of each gene, were considered for further analysis. Sequences with less than 75% of the matrix maximum score were not considered to eliminate a large number of false positives.

DNA binding activity of AvtR: PAGE-EMSA experiments. Eight 39-bp double-stranded DNA fragments (s1 through s8) encompassing the identified 16-bp binding sites were obtained using the corresponding single-stranded oligonucleotides. An unspecific 40-bp double-stranded DNA fragment (5′-CTTATCATTTCATGGATAAGAGGTTCCATGAAA CGCATGG-3′) was used as a negative control. To obtain the labeled double-stranded fragment, one strand of each double-stranded fragment was ^{32}P radiolabeled, followed by annealing with the corresponding reverse oligonucleotide. A 1.5-pmol portion of each double-stranded labeled fragment (~75 ng) was incubated with 2 μl of 50% (vol/vol) glycerin and 0.5 μg of salmon sperm in a total of 20 μl of transcription buffer (TB) for 15 min at 68°C with increasing amounts of AvtR (from 0 to 500 ng). Five microliters of 5× RB loading dye (10 mM Tris-HCl [pH 7.5], 1 mM EDTA, 20% glycerol, 100 μg/ml BSA, 1 mg/ml xylene-cyanol) was added, and the sample was analyzed in a nondenaturing 8% 37.5:1 acrylamide-bisacrylamide gel in TBE buffer. Samples were run for 2 h at 200 V. Gels were vacuum dried, exposed, and developed as described above.

Primer extension analysis. A primer extension protocol was used to define the promoter regions and to precisely determine the transcription start sites of genes *gp29* and *gp30*. A 289-bp DNA fragment covering the shared promoter region between genes *gp29* and *gp30* (positions 15550 to 15839) was generated by PCR and cloned into a pDrive (Qiagen) vector. RNA was produced by incubating 100 ng of the corresponding linear fragment with 6.58 nM *Sulfolobus solfataricus* RNA polymerase, 2.94 nM TATA binding protein (TBP), and 2.86 nM transcription factor B (TFB). Reactions were performed at 70°C for 30 min with 0.2 mM nucleoside triphosphates (NTPs) in 50 μl transcription buffer (50 mM Tris-HCl [pH 8.0], 75 mM KCl, 25 mM $MgCl_2$, and 1 mM DTT). To obtain the gene *gp30* transcription product, AvtR was added to the reaction mixture to a final concentration of 208 μM. The RNA product was annealed to a radiolabeled reverse primer, and either gene *gp29* or *gp30* transcription products were amplified into cDNA in a reverse transcription reaction. The reverse reactions were stopped by addition of 20 μl of 50% formamide loading dye, and 20 μl of the denatured sample was analyzed on a 10% denaturing polyacrylamide gel in TBE buffer. The 5′ termini of the transcripts were mapped by comparison of the primer extension product with a sequencing ladder generated with a Thermo Sequenase cycle sequencing kit (USB Corporation) in a separate reaction with the same primer. Gels were fixed in 10% acetic acid–10% ethanol before vacuum drying, exposure, and development.

IVT. A crenarchaeon-specific *in vitro* transcription (IVT) system was used to test the influence of the purified protein AvtR on transcription from the promoters P*gp29* and P*gp30*, where the *Sulfolobus solfataricus* RNA polymerase and transcription factors TBP and TFB were obtained as described previously (39). A fragment covering the intergenic promoter region between the genes (positions 15550 to 15839) was generated by PCR from genomic AFV6 DNA and cloned directly into a pDrive (Qiagen) cloning vector by T/A cloning. The *Sulfolobus shibatae* virus SSV1 promoter T6 (42) was used as a control in IVT experiments, as previously described (40). *In vitro* transcription reactions were performed using 100 ng of the corresponding plasmid DNA in the presence of 0.2 mM NTPs, 6.58 nM *Sulfolobus solfataricus* RNA polymerase, 2.94 nM TBP, 2.86 nM TFB, and increasing amounts of AvtR (see Fig. 5). The IVT reactions were performed for 30 min at 70°C in 50 μl transcription buffer (50 mM Tris-HCl [pH 8.0], 75 mM KCl, 25 mM $MgCl_2$, and 1 mM DTT). Reactions were stopped by the addition of 250 μl of NEW buffer (10 mM Tris-HCl [pH 8.0], 750 mM NaCl, 10 mM EDTA, 0.5% SDS, and 40 mg/ml glycogen). The newly synthesized *in vitro* RNA was isolated by a phenol-chloroform extraction, followed by ethanol precipitation. Transcription products were detected by primer extension using sequence-specific primers for the previously described viral promoter templates (41, 42). After addition of 20 μl of 50% formamide loading dye, 20 μl of the denatured sample was analyzed on a 10% denaturing polyacrylamide gel in TBE buffer. Gels were fixed in 10% acetic acid–10% ethanol before vacuum drying, exposure, and development.

Oligomerization assays: DNase I footprinting. DNase I footprinting experiments were performed on a 170-bp fragment covering positions 15616 to 15786 of the AFV6 genomic sequence. The two strands of radioactively labeled DNA were generated by PCR in separate reactions by using, in each reaction, one radiolabeled primer. The DNA templates were incubated, in separate reactions, with AvtR in 50 μl transcription buffer for 15 min at 48°C. Samples were then treated for 5 min with 1 U of DNase I (Roche). Reactions were stopped by addition of 200 μl of stop solution (10 mM Tris [pH 8.0], 750 mM NaCl, 10 mM EDTA, 0.5% SDS, 0.04 μg/μl glycogen). DNA fragments were isolated by ethanol precipitation. After addition of 10 μl of 50% formamide loading dye, 5 μl of the denatured samples was analyzed on a denaturing 6% 29:1 acrylamide-bisacrylamide sequencing gel. Samples were run for 2 h at 60 W in TBE buffer. Gels were fixed in 10% acetic acid–10% ethanol before vacuum drying, exposure, and development.

Oligomerization assays: PAGE-EMSA experiments. The mobility of a 289-bp radiolabeled DNA fragment (positions 15550 to 15839) containing the shared promoter region between genes *gp29* and *gp30* was assessed by PAGE-EMSA experiments in the presence of increasing amounts of AvtR. The labeled DNA was generated by PCR by using a previously ^{32}P-radiolabeled forward primer. A 1.5-pmol portion of the double-stranded labeled fragment (~75 ng) was incubated with 2 μl of 50% (vol/vol) glycerin and 0.5 μg of salmon sperm in a total of 20 μl of transcription buffer (TB) for 15 min at 68°C with increasing amounts of AvtR (from 0 to 500 ng). Five microliters of 5× RB loading dye (10 mM Tris-HCl [pH 7.5], 1 mM EDTA, 20% glycerol, 100 μg/ml BSA, 1 mg/ml xylene-cyanol) was added, and the sample was analyzed in a nondenaturing 8% 37.5:1 acrylamide-bisacrylamide gel in TBE buffer. Samples were run for 4 h at 200 V. Gels were vacuum dried, exposed, and developed as described above.

RESULTS

Overall structure of AvtR. Sequence analysis of AvtR suggested the presence of an N-terminal RHH motif most closely related to the CopG protein. CopG contains 45 residues and forms intertwined homodimers. No homologous sequence could be detected for the C-terminal half of AvtR, but fold prediction programs suggested that this part would be structured. We therefore expressed and purified the intact protein. AvtR crystallized in the $P2_1$ space group with 4 copies in the asymmetric unit. The structure was solved at 2.6-Å resolution by SAD phasing using selenomethionine-labeled protein. The statistics on data collection and refinement are provided in Table 1. AvtR is made of two very similar RHH motifs, related by a pseudo-2-fold symmetry (Fig. 1A). β1α1α2 forms the first motif (RHH1, residues 1 to 36) and β3α3α4 the second (RHH2, residues 60 to 100) (Fig. 1F). The linker between the first and the second motifs (24 residues) con-

TABLE 1 Statistics on data collection and structure refinement

Parameter	SeMet value[a]
Space group	$P2_1$
Unit cell (*a*, *b*, *c*) (Å)	50.15, 59.55, 84.04
Resolution (Å)	81.92–2.60 (2.74–2.60)
Total no. of reflections	208,442 (21,911)
Total no. of unique reflections	14,878 (1,989)
Multiplicity	14.0 (11.0)
R_{sym}[a]	0.138 (0.950)
$I/\sigma(I)$	20.9 (3)
Overall completeness (%)	98.8 (91.9)
R/R_{free} (%)[b]	22.10/28.0
RMSD	
Bonds (Å)	0.005
Angles (°)	0.854
Ramachandran plot (%)	
Favored	90
Allowed	10

[a] Values in parentheses indicate the highest resolution shell.
[b] $R_{sym} = \Sigma_h\Sigma_i|I_{hi} - \langle I_h\rangle|/\Sigma_h\Sigma_i I_{hi}$, where I_{hi} is the *i*th observation of the reflection *h* and $\langle I_h\rangle$ is the mean intensity of reflection *h*.
[c] $R = \Sigma||F_o| - |F_c||/|F_o|$. R_{free} was calculated with a small fraction (5%) of randomly selected reflections.

tains a strand (β2) that aligns with β1 and β3 to form the central three-strand antiparallel β-sheet (Fig. 1A). The quasi-2-fold symmetry axis is at the center of the central β-ribbon. The two motifs RHH1 and RHH2 superpose with a root mean square deviation (RMSD) of 2.66 Å for 30 aligned Cα positions. The main difference between the two motifs resides in the length of the β-strands (β3 is 1 residue longer than β1) and the length of the linker between the strand and the helix (the β3-α3 linker contains 5 residues and the β1-α1 linker only 1). Consequently, the β1 strand superposes on the extended β3-α3 linker rather than on strand β3 itself (Fig. 1B). The two motifs share only 13% sequence identity based on the structural alignment.

The RHH motif of AvtR. Ribbon-helix-helix (RHH) DNA binding proteins can use different strategies and oligomeric states for the recognition of their target sites. The structure of the N terminus of AvtR is clearly classified as an RHH motif, which characterizes a large family of transcription factors. Interestingly, the majority of these RHH transcription factors are small proteins (about 45 amino acids), covering one RHH motif, and the functional unit that binds to DNA is created through homodimerization. A search for structural analogues of intact AvtR yielded a good match with the Arc repressor dimer (Protein Data Bank [PDB] code 1myk, RMSD of 2.07 Å for 76 aligned Cα positions). However, the RHH1 motif from AvtR matches best with the CopG monomer (RMSD of 0.79 Å on 35 aligned Cα positions) (Fig. 1C). The best alignment for the RHH2 motif was also obtained with a CopG monomer (RMSD of 2.62 Å for 38 aligned Cα positions) (Fig. 1C). CopG shares 40% sequence identity with AvtR RHH1 but only 11% with AvtR RHH2. Overall, the structures of the CopG dimer and the AvtR monomer also superpose well (RMSD of 2.42 Å) (Fig. 1C): helices from AvtR and CopG occupy identical positions, and the length of the linkers between helices is the same. The β-ribbon of the CopG dimer superposes on the β1 and β3 strands from AvtR. However, the CopG β-strands are longer than those of AvtR.

Identification of an AvtR consensus binding site. The structural analysis of AvtR clearly established that it belongs to the RHH superfamily of DNA binding proteins. As members of this family are most often characterized as or predicted to be transcriptional regulators (6), we decided to confirm the DNA binding properties of AvtR and to identify its targets in the AVF6 genome.

In order to define the hypothetical consensus DNA binding site for AvtR, we used a random binding site selection by the SELEX approach (43, 44; K. R. Nitta, E. Jacox, R. Vincentelli, A. Jolma, D. Sobral, A. Mistral, A. Ohtsuka, A. Kubo, C. Cambillau, Y. Satou, J. Taipale, and P. Lemaire, submitted for publication). The selection of putative AvtR binding sites was achieved using the His-tagged version of the protein (immobilized on Ni^{2+} chelating affinity resin) and a pool of double-stranded DNA oligonucleotides with a central core of 16 random nucleotides (see Materials and Methods). PCR pools from selection rounds 5, 6, and 7 were sequenced by 454 sequencing. The sequence analysis was performed using Bioprospector (45). The consensus binding site was extracted using the Weblogo program (46) and is represented in Fig. 2A.

The random binding site selected by SELEX revealed the presence of a conserved central 10-bp palindromic sequence, GTGGT/ACCAC. Nonetheless, this sequence could not be found in the AFV6 genome. To identify the conserved positions of the AvtR consensus binding site obtained by SELEX, we carried out a systematic mutational analysis of its 16-bp sequence. Out of the 16 analyzed nucleotides, 9 were essential for AvtR binding (Fig. 2B). This information allowed us to refine the data from the SELEX analysis and to better identify the putative consensus AvtR binding site as ATnnTnnTAnnACnTT. Using this site, we were able to screen the AFV6 genome to identify all genes belonging to the putative AvtR regulon.

***In silico* search for AvtR targets.** One way to identify the hypothetical regulon of a predicted transcriptional regulator is to perform *in silico* genomic screenings of the protein's consensus binding site. To identify the putative targets of AvtR, we created a specific PERL script to search for binding sequences in the AFV6 genome using a 16-nucleotide sliding window. The script uses a weighted matrix built by assigning different weights to different nucleotides based on both the data issued from the SELEX experiments and the results of mutational analysis. Only the sequences located within the hypothetical promoter regions, positions −150 to +50 of each gene, and with more than 75% of the matrix maximum score were considered. Eight sites with scores over the cutoff (s1 to s8) were retained for further analysis (Table 2 and Fig. 3A). The eight genes found downstream of the hypothetical targets of AvtR are listed in Table 2.

All predicted binding sites are generally well conserved in the promoter regions of the corresponding homologous genes in the *Betalipothrixvirus* genus of the *Lipothrixviridae* family (Table 3). While sites s1 and s2 are present virtually in all betalipothrixviruses, except AFV7 and AFV9, respectively, other sites, such as s3, s5, s6, s7, and s8, are detected in about half of the members. Site s4, on the other hand, is found only in AFV6 and AFV3. The putative target genes are also well conserved in this viral family. The only exception is gene *gp66*, which is absent from the genomes of AFV7 and SIFV and appears to be duplicated in AFV3. In the genome of the distant relative deltalipothrix virus AFV2, three of the target associated genes, *gp14*, *gp26*, and *gp29*, are also present, even though none of the identified target sites was found in this virus (Table 3).

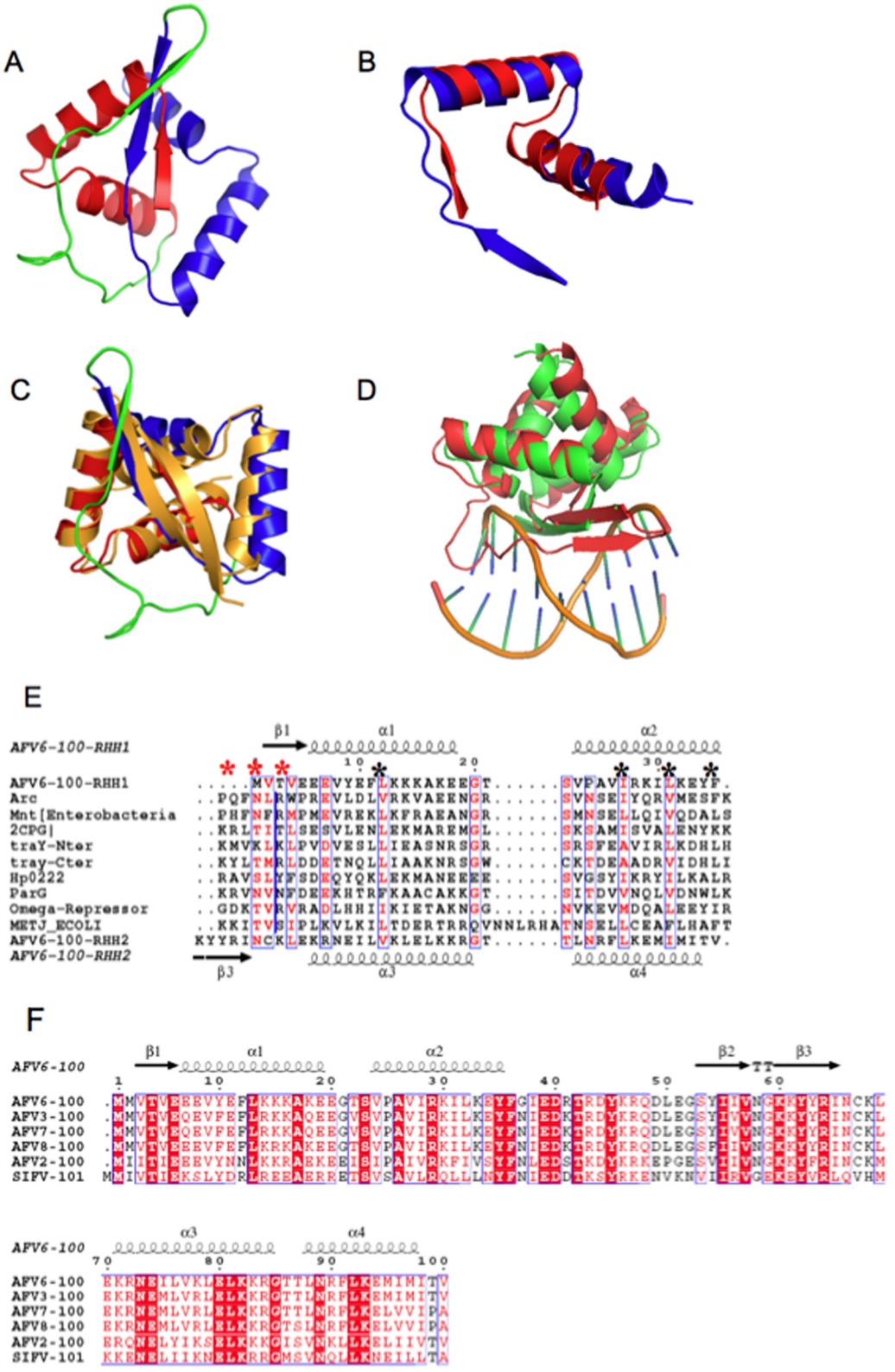

FIG 1 Structure of AvtR. (A) Ribbon presentation of AvtR. The two RHH domains are in red and blue, and the linker containing the extra β-strand is in green. (B) Superposition of the two RHH domains of AvtR. (C) Superposition of AvtR (red, blue, and green) and CopG dimer (gold) (PDB code 1b01). (D) Superposition of the AvtR dimer (red) onto a CopG dimer (green) in complex with a DNA half site (sticks and gold ribbon). (E) Sequence alignments of RHH domains with secondary structure elements as extracted from the AvtR crystal structure. The black asterisks show the conserved hydrophobic positions, and the red asterisks indicate the three amino acids making sequence-specific nucleotide base contacts. (Panels E and F were generated by Espript.) (F) Sequence alignment of AvtR orthologues in lipothrixviruses, with secondary structure elements as extracted from the AvtR crystal structure.

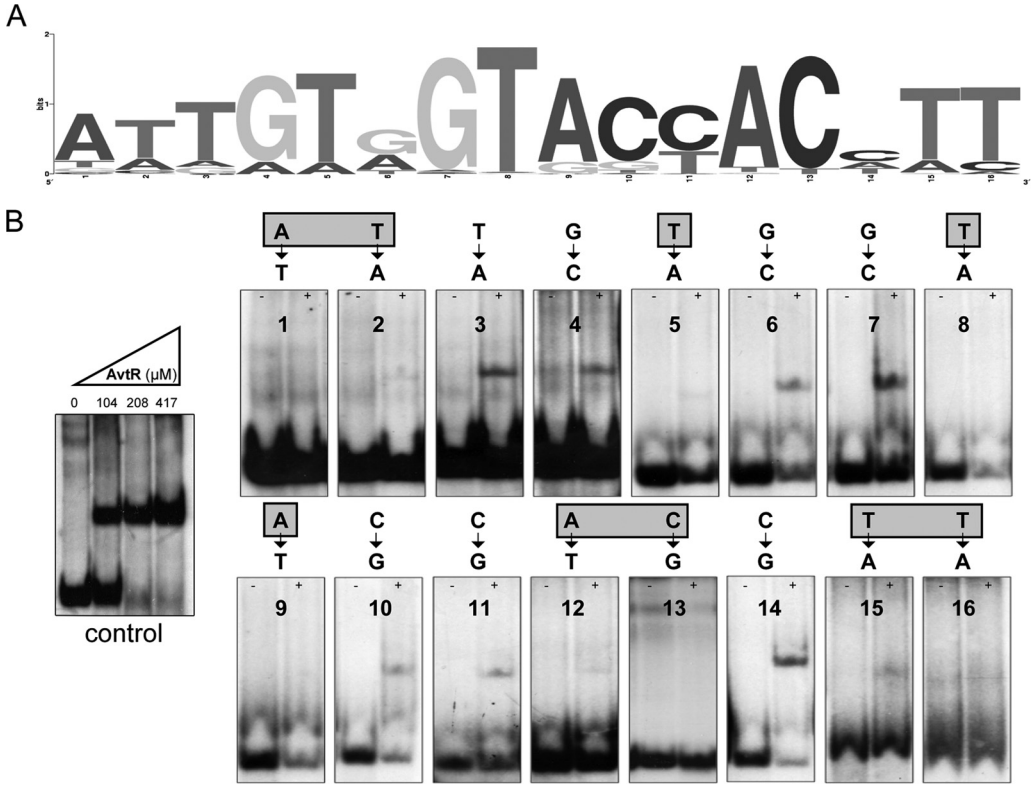

FIG 2 The consensus target site for AvtR and its mutational analysis. (A) SELEX-based prediction of the hypothetical consensus DNA binding site for AvtR. 454 sequencing was done on the PCR pools from selection rounds 5, 6, and 7, the sequence was analyzed using Bioprospector (45), and the consensus binding site was extracted using the Weblogo program. The height of each base is proportional to the percentage of its occurrence in the SELEX experiment at each position. The binding site selected by SELEX shows a central 10-bp palindromic sequence, GTGGT/ACCAC. This sequence could not be found in the AFV6 genome. (B) Mutational analysis of the SELEX predicted AvtR binding sequence. To identify the conserved positions of the AvtR consensus binding site obtained by SELEX, the importance of each nucleotide in the predicted 16-bp binding site was tested by systematic single-nucleotide substitutions. Transversion mutations were performed for each position individually. The same amount of each of the corresponding double-stranded labeled fragments (1.5 pmol) was tested by PAGE-EMSA in the presence of 104 μM AvtR and an excess of nonspecific competitor DNA. Under these conditions, the binding of AvtR to the consensus site is specific. Nucleotides 1, 2, 5, 8, 9, 12, 13, 15, and 16 (gray boxes) were considered to be essential for the site recognition, as mutations in those sites decreased AvtR binding. The remaining substitutions did not alter the DNA binding activity. With this information, the putative consensus AvtR binding site is now defined as ATnnTnnTAnnACnTT.

***In vitro* confirmation of AvtR binding sites.** To verify our *in silico* predictions and the biological significance of the eight putative targets, we assessed the binding of AvtR to these targets *in vitro*. All predicted sites were synthesized as double-stranded 39-bp oligonucleotides containing the target-specific 16-bp sequence. The results of the electrophoretic mobility shift assays for the oligonucleotides representing these sites are presented in Fig. 3B. Six of the eight tested fragments were specifically recognized by AvtR, confirming the results of the *in silico* analysis. No binding was observed for sites s4 and s8 under the conditions used. The biological significance of these two sites remains unclear. These two cases could represent false-positive predictions, although it cannot be excluded that AvtR requires additional factors to recognize these sites.

TABLE 2 Eight putative AvtR targets on the AFV6 genome[a]

Site	Closest gene	Sequence	Position	
			Start	End
s1	*gp11*	TTTATGGTATCACATT	5128	5141
s2	*gp13*	AATGTAGCAGAACAAA	6088	6103
s3	*gp14*	TTTGTGATAGTACATT	6142	6157
s4	*gp16*	TATGTTCTACTACAAT	7209	7224
s5	*gp26*	TTTGTGTTACTACATT	14203	14218
s6	*gp29*	TTTGTAGTGCAATATT	15677	15692
s7	*gp30*	AATGTGCTATCATAAA	15721	15736
s8	*gp66*	AAGGTAGAACAAAAAA	39328	39343

[a] Only the sites located within the hypothetical promoter region, ranging from position −150 to +50, of each gene were considered. Those that had a score of greater than 75% of the matrix maximum value were retained as biologically significant. The start and the end positions within the AFV6 genome sequence are indicated (accession number NC_010152).

A visual inspection of the EMSA results obtained for the eight analyzed sites clearly showed that site s7 is one of the strongest targets for AvtR (Fig. 3B). Site s7 is located in the intergenic region shared between genes *gp29* and *gp30*, which also contains site s6. The two sites are separated by only 43 nucleotides. Interestingly, the *gp30* gene is predicted to code for another putative RHH protein that could also be involved in the gene transcription regulation in AFV6. Also, the presence of a binding site of AvtR upstream of its own gene (*gp29*) suggests that the expression of AvtR is autoregulated. The genetic configuration of this shared promoter region could be a good model to confirm the biological role

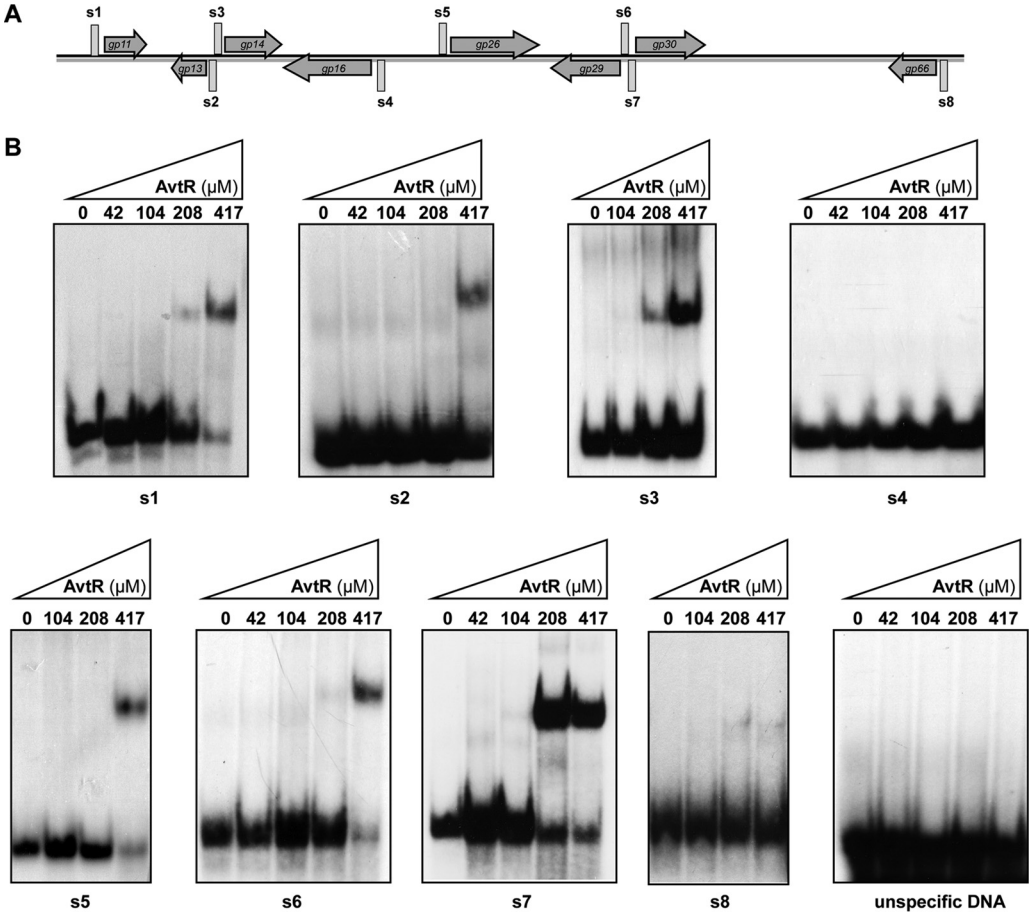

FIG 3 Specific binding of AvtR to the predicted binding sites in the AVF6 genome. (A) AFV6 genome and positions of the predicted AvtR binding sites in the vicinity of viral genes. Sites from s1 to s8 are indicated by vertical lines. The corresponding genes are indicated by arrows. (B) Specific binding of AvtR to six of the eight predicted sites. Short dsDNA oligonucleotides were used to test the AvtR binding. They correspond to the predicted sites s1 to s8 described in Materials and Methods. The PAGE-EMSA tests were performed in the presence of increasing concentrations of AvtR from 0 to 417 μM. Binding to all *in silico*-predicted sites except s4 and s8 is specific. No binding of AvtR is observed with a heterologous DNA fragment of 40 bp. All PAGE-EMSA experiments were performed in the presence of an excess of nonspecific competitor DNA.

of AvtR. Therefore, we decided to use the region s6/s7 for further characterization of AvtR.

Primer extension analysis of the *gp29*/*gp30* promoter region. The transcriptional regulation of the *gp29*/*gp30* genes appears to be complex, as two AvtR binding sites are present in this region. To better understand the transcription regulation of these genes, we decided to determine the positions of the corresponding transcription initiation points by a primer extension analysis. The results of this experiment are presented in Fig. 4. In the absence of AvtR, the *gp30* gene is weakly transcribed. Interestingly, a strong transcription signal was observed in the presence of AvtR in the reaction mixture. This observation of the activation role of AvtR was further confirmed by *in vitro* transcriptional analysis (see next section). The accurate identification of the transcription start site revealed the presence of the typical archaeal conserved sequence elements, the archaeal TATA box and the BRE element, for both genes (Fig. 4B). Even though these sites corresponded to typical archaeal promoters, we were unable to demonstrate the presence of a GTC motif, which is characteristic of the genomes of the rudiviruses SIRV1 and SIRV2 infecting members of the genus *Sulfolobus* (41).

The analysis of the transcription initiation mapping data also highlights that *gp29* and *gp30* transcripts overlap by 45 nucleotides. This could point to a common regulation of their promoters by AvtR. One of the AvtR binding sites is situated 5 nucleotides downstream of the TATA box of *gp29*, whereas the second one is at a distance of 19 nucleotides downstream of the TATA box of *gp30*. On this DNA fragment the AvtR binding sites do not overlap the promoter regions, raising the question of the mechanism of the presumed AvtR transcriptional regulation.

***In vitro* functional analysis of AvtR.** The transcription factors of the RHH superfamily are known to be able to regulate both positively and negatively the transcription of the genes under their control (6). The position of the AvtR binding sites in the promoter regions did not give clear indications about the putative role of AvtR in the transcription of *gp29*/*gp30*. The transcription of these genes was studied using a crenarchaeon-specific *in vitro* transcription system. The transcription activities for these genes were compared in the absence and in the presence of increasing amounts of AvtR. As shown in Fig. 5, AvtR performs a strong repression of the promoter activity of its own gene, *gp29*. The effect is very sharp; in the presence of 100 ng or more of AvtR in solution (417 μM), the

TABLE 3 List of putative AvtR targets in the *Lipothrixviridae* family[a]

Site	Virus	Closest gene	Sequence	Score (max = 16)
s1	AFV6	*gp11*	TTTATGGTATCACATT	16
	AFV2			
	AFV3	*gp11*	TTTATGGTATCACATT	16
	AFV7	*gp06*		
	AFV8	*gp08*	TTTATGGTATCACATT	16
	AFV9	*gp12*	TTTATGGTATCACATT	16
	SIFV	*gp09*	TTTGTACTATCACAAA	11
s2	AFV6	*gp13*	TTTGTTCTGCTACATT	16
	AFV2			
	AFV3	*gp13*	TTTGTTCTACTACATT	15
	AFV7	*gp08*	TTTGTTTTACTACATT	14
	AFV8	*gp10*	TTTGTTCTGCTACATT	16
	AFV9	*gp15*		
	SIFV	*gp12*	TCAATTGTGCTACAAT	12
s3	AFV6	*gp14*	TTTGTGATAGTACATT	16
	AFV2	*gp07*		
	AFV3	*gp14*	TTTGTGATATCATATT	13
	AFV7	*gp09*	TTTGTAGTACTATATT	12
	AFV8	*gp11*	TTTGTGATATCATATT	13
	AFV9			
	SIFV			
s4	AFV6	*gp16*	ATTGTAGTAGAACATA	16
	AFV2			
	AFV3	*gp16*	ATTGTAGTAGAACATA	16
	AFV7	*gp11*		
	AFV8	*gp13*		
	AFV9	*gp25*		
	SIFV	*gp26*		
s5	AFV6	*gp26*	TTTGTGTTACTACATT	16
	AFV2	*gp07*		
	AFV3	*gp25*	TTTGTGTTACTACATT	16
	AFV7	*gp19*	TTTGTGTTACTACATT	16
	AFV8	*gp21*	TTTGTGTTACTACATT	16
	AFV9	*gp23*		
	SIFV	*gp23*		
s6	AFV6	*gp29*	TTTGTAGTGCAATATT	16
	AFV2			
	AFV3	*gp28*	TTTGTAGTGCAATATT	16
	AFV7	*gp22*	TTTGTAGTGCAATATT	16
	AFV8	*gp24*	TTTGTAGTGCAATATT	16
	AFV9	*gp26*		
	SIFV	*gp28*		
s7	AFV6	*gp30*	TTTATGATAGCACATT	16
	AFV2			
	AFV3	*gp29*	TTTATGATAGCACATT	16
	AFV7	*gp23*	TTTATGATAGCACATT	16
	AFV8	*gp25*	TTTATGATAGCACATT	16
	AFV9	*gp27*		
	SIFV	*gp29*		
s8	AFV6	*gp66*	TTTTTTGTTCTACCTT	16
	AFV2			
	AFV3	*gp65*		
	AFV3	*gp68*	TTTTTTGTTCTATCGT	14
	AFV7			
	AFV8	*gp61*	TTTTTTGTTCTATCGT	14
	AFV9	*gp71*	TTTTTTGTTCGATTAC	10
	SIFV			

[a] The identified AvtR targets are present upstream of eight genes in the AFV6 genome. All these genes are conserved and shared in the *Betalipothrix* virus genus of the *Lipothrixviridae* family (AFV3, AFV6, AFV7, AFV8, AFV9, and SIFV). The only exception is gene *gp66*, which is absent from the genomes of AFV7 and SIFV and appears to be duplicated in AFV3. Some of these genes, such as *gp14*, *gp26*, and *gp29*, also could be found in the genome of a more distant relative, the detalipothrixvirus AFV2.

transcription from the *gp29* promoter is totally inhibited. This result clearly shows that AvtR acts as a repressor of the transcription of its own gene. The influence of AvtR on the transcription efficiency of the *gp30* gene is more complex. Even though in the absence of AvtR, this gene is weakly transcribed (Fig. 5B), in the presence of a low concentration of AvtR (208 μM), transcription from *gp30* is strongly activated. Unexpectedly, at higher concentrations of the regulator, the activation effect disappears progressively. If the activation effect obtained for the transcription of gene *gp30* at a relatively low AvtR concentration most probably reflects the situation *in vivo*, the subsequent repression seen with higher protein concentrations could either have biological significance or be an *in vitro* artifact. No influence of AvtR on transcription when using a control heterologous promoter, *T6* from SSV1, has been detected.

In conclusion, at the lowest tested concentration, AvtR's function in transcription regulation is clearly target dependent. The protein acts as a transcription repressor for P*gp29* but is able to activate the transcription from the promoter P*gp30*. Interestingly, at higher concentrations AvtR continues to represses transcription from its own gene (*gp29*) and no longer activates P*gp30*.

Oligomerization of AvtR on a DNA template. To better characterize the regulation of transcription exerted by AvtR, we performed a DNase I footprint assay on the DNA region directly involved in the interaction between the protein and the promoters of genes *gp29* and *gp30* under its control.

As seen in Fig. 6A, large protected regions were identified on both DNA strands using increasing concentrations of the protein. AvtR protects from DNase I digestion a region of about 100 nt situated between the TATA boxes of genes *gp29* and *gp30*. Being a relatively small protein of only 11.98 kDa, AvtR could protect a DNA fragment of that size only if several subunits of the protein form an oligomer tightly protecting the DNA target. The initiation of this oligomerization is certainly dependent on the presence of the primary binding sites, s6 and s7. Despite the fact that high concentrations of AvtR in the probe appear to be required in order to achieve this oligomerization, indicating that this process is not very efficient *in vitro*, this process still appears to be specific. Even at the highest concentration of AvtR used, the regions following both TATA boxes remain unprotected (lower parts of gels A1 and A2 in Fig. 6). It remains unclear why the oligomerization does not include the loci containing the TATA boxes.

To independently verify the occurrence of the oligomerization of AvtR on its target DNA, the migration of a 289-bp DNA fragment encompassing the shared promoter region between genes *gp29* and *gp30* was analyzed by PAGE-EMSA in the presence of increasing concentrations of AvtR. We observed a gradual decrease of the fragment's mobility in the presence of increasing concentrations of AvtR (Fig. 6B). This result confirms the initial oligomerization hypothesis, which is further presented in Discussion and schematically summarized in Fig. 6C.

DISCUSSION

The genomes of archaeal viruses are remarkably rich in RHH proteins. A recent *in silico* analysis of 10 crenarchaeal virus genomes revealed the presence of at least 18 putative RHH regulators in the analyzed genomes (2). Their role in the regulation of the virus infectious cycle remains largely unknown. The only examples of well-studied RHHs in an archaeal context are protein ORF56, encoded by the *Sulfolobus islandicus* plasmid pRN1 (22, 25, 47–49),

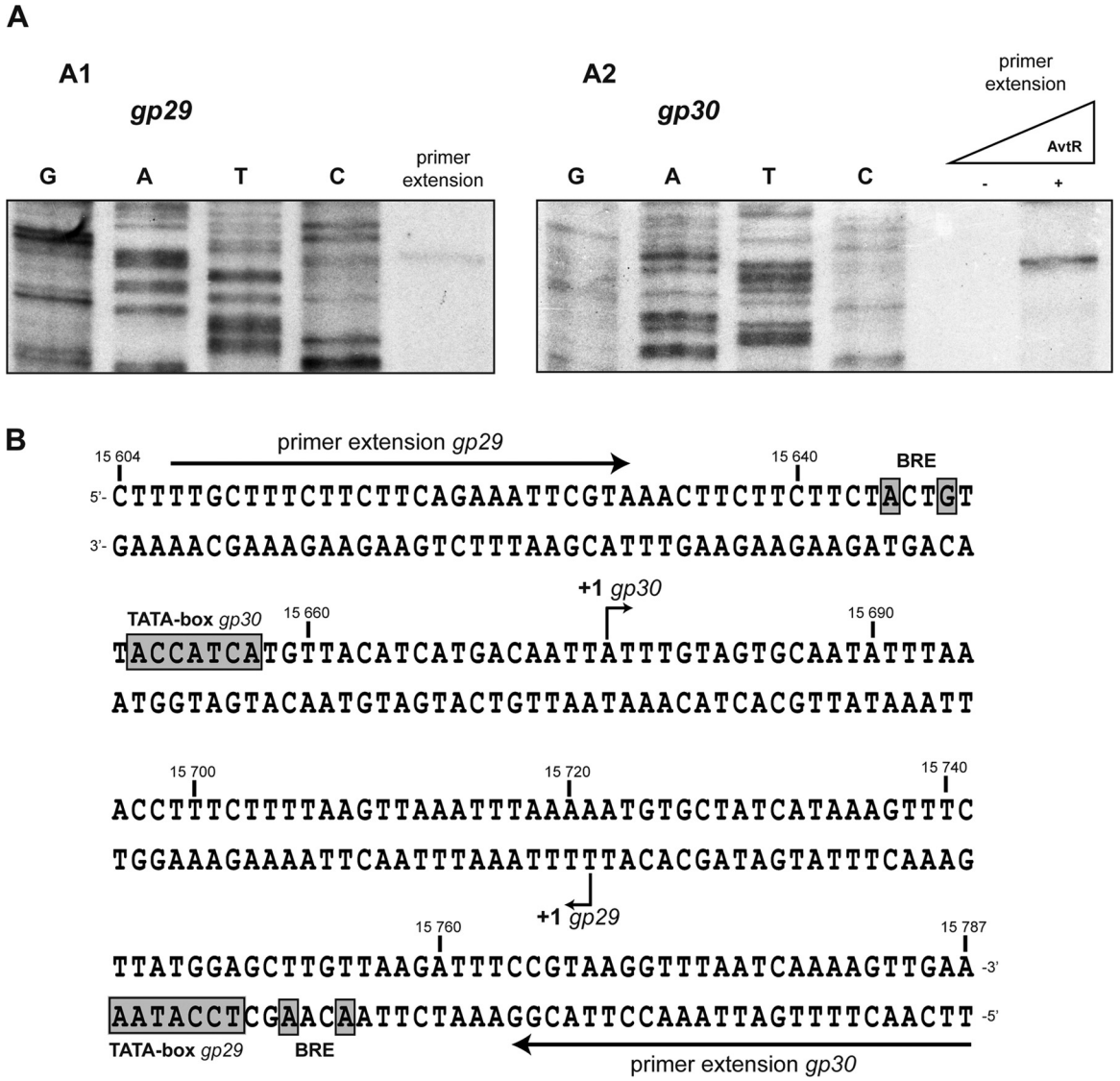

FIG 4 Mapping of the transcription initiation sites for *gp29* and *gp30* of AFV6. (A) Primer extension assays. The positions of the 5′ termini of each *in vitro* generated transcript were mapped using the sequence of the 289-bp fragment covering the *gp29/gp30* intergenic region. Identification of the +1 position for the *gp29* (A1) and for the *gp30* (A2) genes is shown. The transcript for the *gp30* gene was obtained in the presence of AvtR. (B) Detailed map of transcription signals present in the *gp29/gp30* region of AFV6. Only 183 out of 289 nucleotides are represented. The positions of the transcription initiations points (+1), the BRE sites, and the TATA boxes for both genes, *gp29* and *gp30*, are indicated, as well as the positions of primers used for the extension and sequencing reactions.

the regulator SvtR, encoded by the rudivirus SIRV (5), and the recently described protein E73, encoded by the virus SSV-RH (22). ORF56 regulates the plasmid copy number, SvtR is involved in the regulation of the lytic cycle of the rudiviruses, and E73 has been suggested to target elements of the host genome. They exist as dimers in solution and bind to their cognate DNAs as pairs of dimers, without any apparent cooperativity.

In this study, we comprehensively characterized the structure, function, and DNA targets of AvtR (*Acidianus* virus transcription regulator), an RHH protein encoded by the virus AFV6, infecting the hyperthermophilic crenarchaeon *Acidianus.* In the absence of their cognate DNAs, RHH proteins form symmetric dimers with a 2-fold symmetry axis centered on the two-stranded antiparallel β-sheet. In contrast with other well-characterized RHH proteins, the AvtR protein is composed of two tandem RHH motifs that are connected by a linker that aligns with strands β1 and β3 to form a central three-stranded antiparallel β-sheet. Orthologues of AvtR are present and well conserved in several members of the *Lipothrixviridae* viral family (Fig. 1F). The RHHs of AvtR are arranged as a pseudosymmetric dimer, superposable to dimeric RHH proteins. The TraY protein from the *E. coli* episome F is also an RHH tandem repeat (50), but its structure is unknown. The MetJ (24), Arc (21), Ω (18), and CopG (11, 27–29) RHH proteins all bind to their DNA recognition sites as dimers of dimers.

Sequence comparison shows that RHH motifs have no absolutely conserved amino acid positions but that they share certain sequence features: (i) the β-ribbons present a pattern of alternating hydrophilic/hydrophobic side chains, (ii) a G-X-T/S/N motif is conserved in the loop between helix α1 and helix α2 (except in MetJ), and (iii) four positions in helix α1 and α2 are conserved as

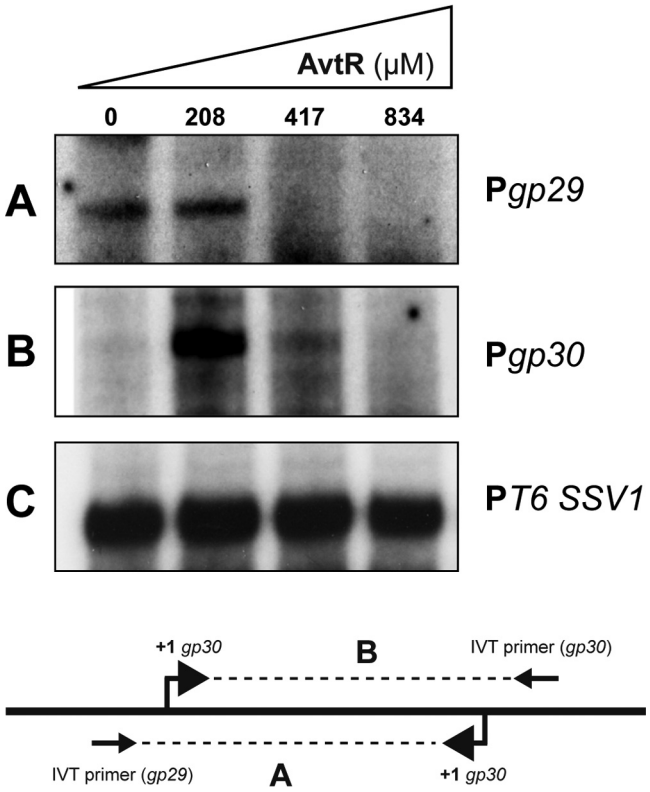

FIG 5 AvtR regulates transcription from the promoters of the *gp29* and *gp30* genes. The involvement of AvtR in the transcription regulation of *gp29* and *gp30* was demonstrated using a host reconstituted *in vitro* transcription system (IVT). All IVT reactions were performed in the presence of limiting amounts of transcription factors TBP and TFB. (A and B) Gene-specific transcription assays were performed on the promoters of AFV6 genes *gp29* (A) and *gp30* (B). The concentrations of AvtR in each reaction mixture are indicated in μM. Bottom, schematic representation of the positions of the primers used for the IVT reactions for *gp29* and *gp30* and of the respective transcription start sites. AvtR strongly represses the transcription of its own gene, *gp29*, and represses that of the *gp30* promoter regulation. (C) To demonstrate that the AvtR activity is target specific, the IVT reaction was performed under the same conditions as for panels A and B using a T6 promoter of the unrelated virus SSV1.

hydrophobic (black asterisks in Fig. 1E). The pattern of hydrophilic/hydrophobic residues along the β-strand, a hallmark of RHH proteins, is present in AvtR along β1 (RHH1) and along the linker between strand β3 and helix α2 (RHH2). Although this linker does not form a regular β-strand, it is an integral part of the ribbon of the RHH motif (Fig. 1C). As observed in structures of RHH-DNA oligonucleotide complexes, the central β-sheet of the dimer specifically recognizes bases from the major groove of the DNA target sequence: three amino acid side chains from each strand (corresponding to Arg4, Thr6, and Thr8 in the CopG/DNA structure [red asterisks in Fig. 1D]) point into the major DNA groove and make crucial sequence-specific nucleotide contacts. Amino acid residues located at these positions define the DNA binding specificity of the RHH regulators. Although there is no strict conserved sequence pattern, a basic residue (Arg or Lys) is always present at position 1 or 3. Based on the position of this basic residue, RHH proteins can be separated into two classes (10). Class I proteins have a basic residue at the start of the β-strand (CopG, ParD, ParG, and MetJ) whereas class II proteins have a basic residue at the end (Arc, omega, and Mnt repressors). RHH proteins usually bind DNA as dimers contacting two or more inverted or tandem repeats within their operators and form higher-order oligomers (6, 11, 27, 51, 52). It should be noted that the two strands of the central β-sheet in RHH dimers do not interact symmetrically with the cognate DNA. The arrangement of the central β-sheet in AvtR is different from that observed for other RHH proteins: the β1 strand (RHH1) is much shorter than the homologous strands of other RHH members and does not contain basic residues, while strand β3 (RHH2) contains arginine residues at its N and C termini. However, the main difference from canonical RHH structures resides in the presence of a third strand in the central sheet in AvtR. Superposition of the structure of the AvtR dimer onto that of the CopG-DNA complex shows that it cannot bind to its cognate DNA in the same way, because the third strand clashes with the phosphate backbone (Fig. 1D). In order to bind its DNA recognition site, AvtR must either rearrange the linker strand to relieve the clash or use a noncanonical DNA binding mode.

All RHH proteins studied in detail so far are involved in transcription regulation. They recognize specific DNA binding sites and influence the efficiency of transcription initiation. The AvtR binding site as obtained by the SELEX methodology was not found in the sequence of the AFV6 genome. A systematic mutational analysis of the conserved nucleotides allowed us to predict *in silico* 8 putative AvtR binding sites in the AFV6 genome. At least 6 of these sites could be *in vivo* targets for AvtR, as specific binding of the protein to these targets was demonstrated *in vitro*. The other two remaining sites either could have been wrongly predicted or might require additional factors to be recognized by AvtR.

The identified binding sites are generally well conserved in the *Lipothrixviridae* viral family (Table 3). Viruses AFV6 and AFV3 appear to be the closest relatives in this family, as all identified sites are always present in both genomes. In the more distantly related deltalipothrixvirus AFV2, none of the target sites was found, even though three of the identified target-associated genes were present in its genome sequence. Most of the genes under AvtR control are also well conserved throughout this family and could play an important role in these virus cycles. For example, some of these genes code for structural proteins conserved among all betalipothrixviruses (*gp11* and *gp66*). Also, the protein encoded by gene *gp16* (annotated as a DNA binding protein) is present in several crenarchaeal viruses, such as betalipothrixviruses, the rudiviruses, the icosahedral virus STIV1, the unclassified STIV2, the bicaudavirus ATV, and the gammalipothrixvirus AFV1 (53). For the remaining genes, no predictions could be made for the proteins they code for, and the complete biological scenario behind the regulation of AvtR in the AVF6 viral cycle remains unclear.

One of the regions recognized by AvtR in the AFV6 genome is particularly interesting in its genetic organization and complexity of its transcription regulation. This region includes the overlapping promoters of genes *gp29* (AvtR's own gene) and *gp30*, coding for another hypothetical RHH protein. We showed here that AvtR strongly represses transcription from the promoter of its own gene, *gp29*, but activates the promoter of *gp30*. The position of the AvtR binding sites in the studied intergenic region raises the question about the mechanism of regulation. The primary AvtR binding sites do not overlap the promoters of *gp29* and *gp30*. DNase I footprint analysis and mobility shift assays performed with DNA fragments carrying both binding sites provide strong indications that AvtR is capable of oligomerization on the DNA template,

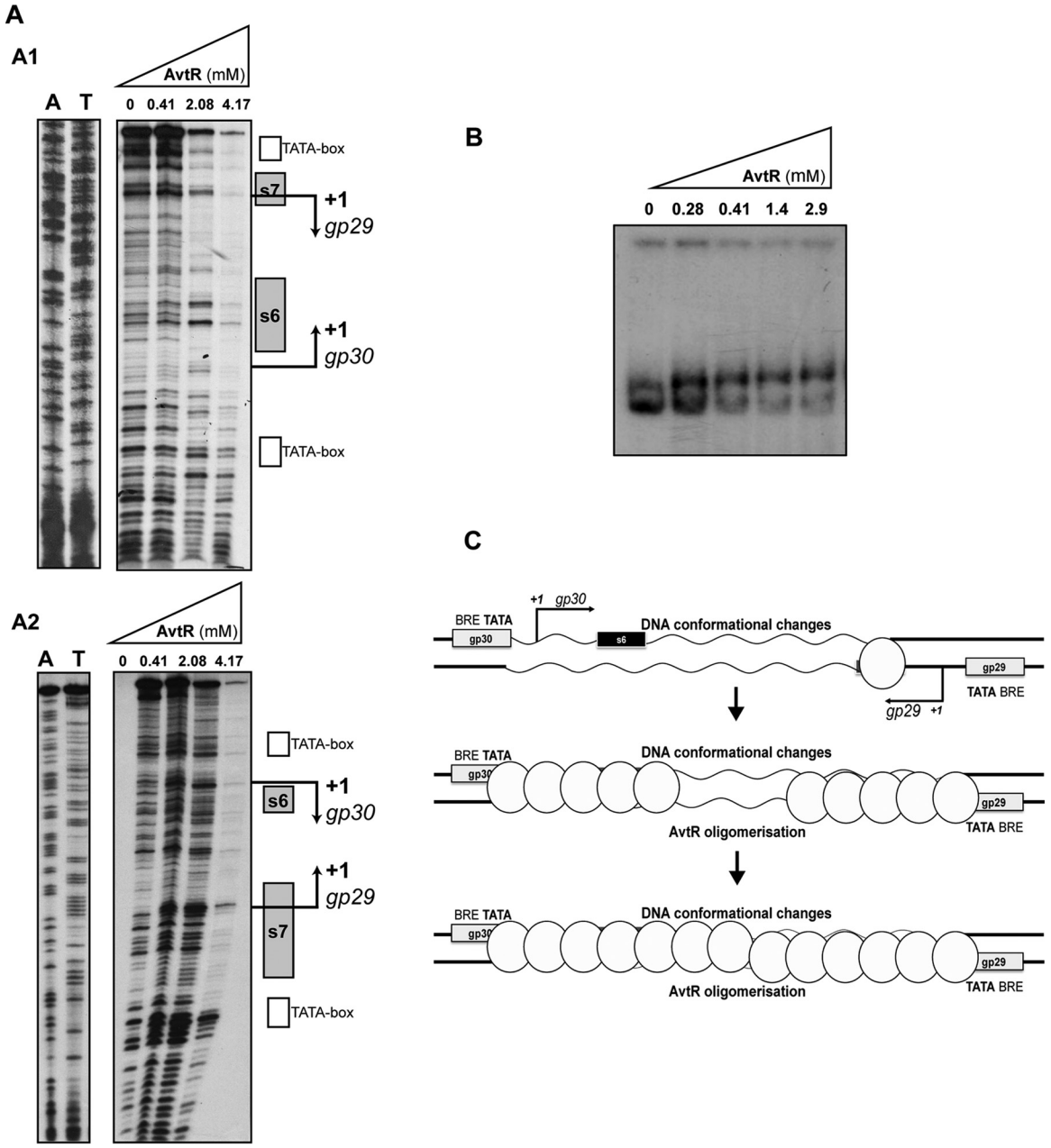

FIG 6 DNase I footprinting analysis of the regions protected by AvtR in the *gp29* and *gp30* promoters. (A) DNase I footprinting assays performed on a 170-bp fragment corresponding to the *gp29*/*gp30* intergenic region. Leading and lagging DNA strands were analyzed separately (A1 and A2, respectively), and the concentrations of AvtR present prior to the addition of DNase I are indicated on the upper part of each gel. The gray boxes represent the identified DNA binding sites of AvtR, and the arrows indicate the transcription start points and their directions. The AvtR oligomerization starts from both of its binding sites and extends upstream and downstream to a total of 97 nucleotides. (B) Migration of a 289-bp DNA fragment containing the promoter region between genes *gp29* and *gp30*, analyzed by PAGE-EMSA. A gradual decrease in the mobility of the fragment is seen in the presence of increasing concentrations of AvtR, confirming the oligomerization hypothesis. (C) Proposed model of regulation by AvtR. The shared promoter region of genes *gp29* and *gp30* contains two binding sites for AvtR. When present at a low concentration, AvtR binds primarily to site s7. This binding activates transcription from gene *gp30*, probably by inducing conformational changes on the promoter DNA (shown by wavy line). At higher concentrations, AvtR occupies both sites s6 and s7. Oligomerization of AvtR will then cover both promoter's TATA boxes and impair the formation of the transcription initiation complex, repressing transcription from both genes *gp29* and *gp30*.

most likely starting from the primary sites s6 and s7 and subsequently covering the entire region between the TATA boxes of *gp29* and *gp30*. An analogous situation is described for the transcriptional repressor CopG (28). Initial binding of the CopG dimer to its specific site takes place through the major groove of DNA. At higher repressor concentrations, binding of CopG to the primary site may promote successive binding to secondary sites in the operator (27). We propose a similar mechanism to explain how a small protein such as AvtR could protect the long intergenic *gp29*/*gp30* region. Other archaeal regulatory proteins with different DNA binding motifs, such as TrpY (54) and FL11 (55), are also capable of binding to control regions as multiple copies in a con-

centration-dependent manner. In this case, the oligomerization of AvtR appears to be guided by short degenerate secondary sites. A careful analysis of the sequence located between the two initial binding sites, s6 and s7, revealed the presence of several short repeats (TTTAA) that could play this role.

A hypothetical model integrating all observed events is presented in Fig. 6C. At low concentrations, AvtR binds only to the strongest site, s7, situated 62 nucleotides downstream of the TATA box of *gp30*. We presume that this binding causes the activation of *gp30* transcription. At this stage, no exact mechanism for this effect can be proposed. One hypothesis could be that this binding induces some conformational changes in the promoter region. At higher concentrations, AvtR will also bind to the weaker second site, s6, and the oligomerization starting from both sites s7 and s6 could cover the *gp29/gp30* intergenic region, reach the TATA boxes, and then directly interact with the transcription initiation complexes. The biological significance of this unusual mechanism of regulation remains unclear. One could suppose that the intracellular concentration of AvtR is a crucial parameter of the system that plays a role of the "molecular timer" activating and repressing viral genes depending on the stage of the virus development cycle.

To conclude, AvtR represents the second example of an RHH protein encoded by archaeal viruses, and it is distinguished from other RHH proteins by some unusual structural and regulatory properties. The relatively high number of genes under AvtR control clearly indicates the importance of this regulator in the AFV6 infection cycle.

ACKNOWLEDGMENTS

This work was supported by a VIRAR grant (NT05-2_41674) from the Agence Nationale de Recherche. N.P. was supported by the EU Marie Curie research training network SOLAR (grant MRTN-CT-2006-033499).

We thank Chloé Danioux for help and useful discussions.

REFERENCES

1. **Pina M, Bize A, Forterre P, Prangishvili D.** 2011. The archeoviruses. FEMS Microbiol. Rev. **35**:1035–1054.
2. **Prangishvili D, Garrett RA, Koonin EV.** 2006. Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. Virus Res. **117**:52–67.
3. **Geslin C, Gaillard M, Flament D, Rouault K, Le Romancer M, Prieur D, Erauso G.** 2007. Analysis of the first genome of a hyperthermophilic marine virus-like particle, PAV1, isolated from *Pyrococcus abyssi.* J. Bacteriol. **189**:4510–4519.
4. **Krupovic M, White MF, Forterre P, Prangishvili D.** 2012. Postcards from the edge: structural genomics of archaeal viruses. Adv. Virus Res. **82**:33–62.
5. **Guilliere F, Peixeiro N, Kessler A, Raynal B, Desnoues N, Keller J, Delepierre M, Prangishvili D, Sezonov G, Guijarro JI.** 2009. Structure, function, and targets of the transcriptional regulator SvtR from the hyperthermophilic archaeal virus SIRV1. J. Biol. Chem. **284**:22222–22237.
6. **Schreiter ER, Drennan CL.** 2007. Ribbon-helix-helix transcription factors: variations on a theme. Nat. Rev. Microbiol. **5**:710–720.
7. **Burgering MJ, Boelens R, Gilbert DE, Breg JN, Knight KL, Sauer RT, Kaptein R.** 1994. Solution structure of dimeric Mnt repressor (1-76). Biochemistry **33**:15036–15045.
8. **Chopra N, Agarwal S, Verma S, Bhatnagar S, Bhatnagar R.** 2011. Modeling of the structure and interactions of the *B. anthracis* antitoxin, MoxX: deletion mutant studies highlight its modular structure and repressor function. J. Comp. Aid. Mol. Design **25**:275–291.
9. **Gallo M, Ferrari E, Eliseo T, Amata I, Pertinhez T, Katsuyama A, Paci M, Farah C, Spisni A, Cicero D.** 2010. A new member of the ribbon-helix-helix transcription factor superfamily from the plant pathogen *Xanthomonas axonopodis* pv.citri. J. Struct. Biol. **170**:21–31.
10. **Golovanov AP, Barilla D, Golovanova M, Hayes F, Lian LY.** 2003. ParG, a protein required for active partition of bacterial plasmids, has a dimeric ribbon-helix-helix structure. Mol. Microbiol. **50**:1141–1153.
11. **Gomis-Ruth FX, Sola M, Acebo P, Parraga A, Guasch A, Eritja R, Gonzalez A, Espinosa M, del Solar G, Coll M.** 1998. The structure of plasmid-encoded transcriptional repressor CopG unliganded and bound to its operator. EMBO J. **17**:7404–7415.
12. **Huang L, Yin P, Zhu X, Zhang Y, Ye K.** 2011. Crystal structure and centromere binding of the plasmid segregation protein ParB from pCXC100. Nucleic Acids Res. **39**:2954–2968.
13. **Larson JD, Jenkins JL, Schuermann JP, Zhou Y, Becker DF, Tanner JJ.** 2006. Crystal structures of the DNA-binding domain of *Escherichia coli* proline utilization A flavoprotein and analysis of the role of Lys9 in DNA recognition. Protein Sci. **15**:2630–2641.
14. **Li G-Y, Zhang Y, Inouye M, Ikura M.** 2008. Structural mechanism of transcriptional autorepression of the *Escherichia coli* RelB/RelE antitoxin/toxin module. J. Mol. Biol. **380**:107–119.
15. **Lu J, den Dulk-Ras A, Hooykaas P, Glover M.** 2009. *Agrobacterium tumefaciens* VirC2 enhances T-DNA transfer and virulence through its C-terminal ribbon-helix-helix DNA-binding fold. Proc. Natl. Acad. Sci. U. S. A. **106**:9643–9648.
16. **Madl T, Van Melderen L, Mine N, Respondek M, Oberer M, Keller W, Khatai L, Zangger K.** 2006. Structural basis for nucleic acid and toxin recognition of the bacterial antitoxin CcdA. J. Mol. Biol. **364**:170–185.
17. **Mattison K, Wilbur JS, So M, Brennan RG.** 2006. Structure of FitAB from *Neisseria gonorrhoeae* bound to DNA reveals a tetramer of toxin-antitoxin heterodimers containing pin domains and ribbon-helix-helix motifs. J. Biol. Chem. **281**:37942–37951.
18. **Murayama K, Orth P, de la Hoz AB, Alonso JC, Saenger W.** 2001. Crystal structure of omega transcriptional repressor encoded by *Streptococcus pyogenes* plasmid pSM19035 at 1.5 A resolution. J. Mol. Biol. **314**:789–796.
19. **Popescu A, Karpay A, Israel DA, Peek RM, Jr, Krezel AM.** 2005. *Helicobacter pylori* protein HP0222 belongs to Arc/MetJ family of transcriptional regulators. Proteins **59**:303–311.
20. **Pryor E, Waligora E, Xu B, Dellos-Nolan S, Wozniak D, Hollis T.** 2012. The transcription factor AmrZ utilizes multiple DNA binding modes to recognize activator and repressor sequences of *Pseudomonas aeruginosa* virulence genes. PLoS Pathog. **8**:e1002648. doi:10.1371/journal.ppat.1002648.
21. **Raumann BE, Rould MA, Pabo CO, Sauer RT.** 1994. DNA recognition by beta-sheets in the Arc repressor-operator crystal structure. Nature **367**:754–757.
22. **Schlenker C, Goel A, Tripet BP, Menon S, Willi T, Dlakic M, Young MJ, Lawrence CM, Copie V.** 2012. Structural studies of E73 from a hyperthermophilic archaeal virus identify the "RH3" domain, an elaborated ribbon-helix-helix motif involved in DNA recognition. Biochemistry **51**:2899–2910.
23. **Schreiter ER, Wang SC, Zamble DB, Drennan CL.** 2006. NikR-operator complex structure and the mechanism of repressor activation by metal ions. Proc. Natl. Acad. Sci. U. S. A. **103**:13676–13681.
24. **Somers WS, Phillips SE.** 1992. Crystal structure of the met repressor-operator complex at 2.8 A resolution reveals DNA recognition by beta-strands. Nature **359**:387–393.
25. **Weininger U, Zeeb M, Neumann P, Löw C, Stubbs M, Lipps G, Balbach J.** 2009. Structure-based stability analysis of an extremely stable dimeric DNA binding protein from *Sulfolobus islandicus.* Biochemistry **48**:10030–10037.
26. **Wong J, Lu J, Edwards R, Frost L, Glover M.** 2011. Structural basis of cooperative DNA recognition by the plasmid conjugation factor, TraM. Nucleic Acids Res. **39**:6775–6788.
27. **Costa M, Sola M, del Solar G, Eritja R, Hernandez-Arriaga AM, Espinosa M, Gomis-Ruth FX, Coll M.** 2001. Plasmid transcriptional repressor CopG oligomerises to render helical superstructures unbound and in complexes with oligonucleotides. J. Mol. Biol. **310**:403–417.
28. **del Solar G, Hernandez-Arriaga AM, Gomis-Ruth FX, Coll M, Espinosa M.** 2002. A genetically economical family of plasmid-encoded transcriptional repressors involved in control of plasmid copy number. J. Bacteriol. **184**:4943–4951.
29. **Hernandez-Arriaga AM, Rubio-Lepe TS, Espinosa M, del Solar G.** 2009. Repressor CopG prevents access of RNA polymerase to promoter and actively dissociates open complexes. Nucleic Acids Res. **37**:4799–4811.
30. **He YY, McNally T, Manfield I, Navratil O, Old IG, Phillips SE, Saint-**

Girons I, Stockley PG. 1992. Probing met repressor-operator recognition in solution. Nature **359**:431–433.
31. **Vestergaard G, Aramayo R, Basta T, Haring M, Peng X, Brugger K, Chen L, Rachel R, Boisset N, Garrett RA, Prangishvili D.** 2008. Structure of the acidianus filamentous virus 3 and comparative genomics of related archaeal lipothrixviruses. J. Virol. **82**:371–381.
32. **Leslie AGW.** 1992. Joint CCP4 and EACMB Newsl. Protein Crystallogr., no. 26.
33. **Evans P.** 2006. Scaling and assessment of data quality. Acta Crystallogr. D Biol. Crystallogr. **62**:72–82.
34. **Grosse-Kunstleve RW, Adams PD.** 2003. Substructure search procedures for macromolecular structures. Acta Crystallogr. D Biol. Crystallogr. **59**: 1966–1973.
35. **Bricogne G, Vonrhein C, Flensburg C, Schiltz M, Paciorek W.** 2003. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. Acta Crystallogr. D Biol. Crystallogr. **59**:2023–2030.
36. **Terwilliger TC.** 2000. Maximum-likelihood density modification. Acta Crystallogr. D Biol. Crystallogr. **56**:965–972.
37. **Jones TA, Kjeldgaard M.** 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A **47**:110–119.
38. **Murshudov GN, Vagin AA, Dodson EJ.** 1997. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. **53**:240–255.
39. **Abella M, Rodriguez S, Paytubi S, Campoy S, White MF, Barbe J.** 2007. The *Sulfolobus solfataricus* radA paralogue sso0777 is DNA damage inducible and positively regulated by the Sta1 protein. Nucleic Acids Res. **35**: 6788–6797.
40. **Kessler A, Sezonov G, Guijarro JI, Desnoues N, Rose T, Delepierre M, Bell SD, Prangishvili D.** 2006. A novel archaeal regulatory protein, Sta1, activates transcription from viral promoters. Nucleic Acids Res. **34**:4837–4845.
41. **Kessler A, Brinkman AB, van der Oost J, Prangishvili D.** 2004. Transcription of the rod-shaped viruses SIRV1 and SIRV2 of the hyperthermophilic archaeon *Sulfolobus.* J. Bacteriol. **186**:7745–7753.
42. **Qureshi SA, Bell SD, Jackson SP.** 1997. Factor requirements for transcription in the Archaeon *Sulfolobus shibatae.* EMBO J. **16**:2927–2936.
43. **Jolma A, Kivioja T, Toivonen J, Cheng L, Wei G, Enge M, Taipale M, Vaquerizas JM, Yan J, Sillanpaa MJ, Bonke M, Palin K, Talukder S, Hughes TR, Luscombe NM, Ukkonen E, Taipale J.** 2010. Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. Genome Res. **20**:861–873.
44. **Roulet E, Busso S, Camargo AA, Simpson AJ, Mermod N, Bucher P.** 2002. High-throughput SELEX SAGE method for quantitative modeling of transcription-factor binding sites. Nat. Biotechnol. **20**:831–835.
45. **Liu X, Brutlag DL, Liu JS.** 2001. BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. Pac. Symp. Biocomput. **6**:127–138.
46. **Crooks GE, Hon G, Chandonia JM, Brenner SE.** 2004. WebLogo: a sequence logo generator. Genome Res. **14**:1188–1190.
47. **Berkner S, Lipps G.** 2007. Characterization of the transcriptional activity of the cryptic plasmid pRN1 from *Sulfolobus islandicus* REN1H1 and regulation of its replication operon. J. Bacteriol. **189**:1711–1721.
48. **Lipps G.** 2004. The replication protein of the *Sulfolobus islandicus* plasmid pRN1. Biochem. Soc. Trans. **32**:240–244.
49. **Lipps G, Stegert M, Krauss G.** 2001. Thermostable and site-specific DNA binding of the gene product ORF56 from the *Sulfolobus islandicus* plasmid pRN1, a putative archael plasmid copy control protein. Nucleic Acids Res. **29**:904–913.
50. **Lum PL, Rodgers ME, Schildbach JF.** 2002. TraY DNA recognition of its two F factor binding sites. J. Mol. Biol. **321**:563–578.
51. **Chivers PT, Sauer RT.** 2000. Regulation of high affinity nickel uptake in bacteria. Ni2+-dependent interaction of NikR with wild-type and mutant operator sites. J. Biol. Chem. **275**:19735–19741.
52. **Saint-Girons I, Duchange N, Cohen GN, Zakin MM.** 1984. Structure and autoregulation of the metJ regulatory gene in *Escherichia coli.* J. Biol. Chem. **259**:14282–14285.
53. **Larson ET, Eilers BJ, Reiter D, Ortmann AC, Young MJ, Lawrence CM.** 2007. A new DNA binding protein highly conserved in diverse crenarchaeal viruses. Virology **363**:387–396.
54. **Karr EA, Sandman K, Lurz R, Reeve JN.** 2008. TrpY regulation of trpB2 transcription in *Methanothermobacter thermautotrophicus.* J. Bacteriol. **190**:2637–2641.
55. **Yokoyama K, Ishijima SA, Koike H, Kurihara C, Shimowasa A, Kabasawa M, Kawashima T, Suzuki M.** 2007. Feast/famine regulation by transcription factor FL11 for the survival of the hyperthermophilic archaeon *Pyrococcus* OT3. Structure **15**:1542–1554.